













The all reise a company of the second compan











































**Research Data Life Cycle** CIHR IRSO CRSNG SSHRC≡CRSH **Repository Storage** re a.org **REGISTRY OF RESEARCH DATA REPOSITORIES Discovery & Access** FRDR DFDR DISSEMINATE borealis The Canadian Dataverse Repo lunaris PLAN CREATE PROCESS ANALYZE PRESERVE REUSE INDEX































































**Managing Your Data** Large ... ... . . . The 'long-tail' of Computing Medium . Needs research data... Small Medium Large Small Size of Data 







































































<image>

































































#### ChatGPT

## What are some of the challenges of managing data on high performance computing platforms?

![](_page_63_Picture_5.jpeg)

- **Data transfer and storage**: Moving large amounts of data to and from HPC systems can be time-consuming and costly. Additionally, storing data on HPC systems can be challenging due to limited storage capacity and the need for high-speed access.
- **Data management**: Managing and organizing large datasets on HPC systems can be difficult, as traditional file systems may not be well-suited for the high-performance needs of HPC applications.
- **Data integrity and security**: Ensuring the integrity and security of data on HPC systems can be challenging, as these systems are often shared by multiple users and may be exposed to a higher risk of data breaches.
- **Data visualization**: Visualizing and analyzing large, complex datasets on HPC systems can be difficult, as traditional visualization tools may not be able to handle the scale and complexity of the data.
- **Scalability:** Ensuring that data management solutions can scale to meet the needs of HPC systems as they grow in size and complexity can be a major challenge.

### ChatGPT

# What are some of the challenges of managing data on high performance computing platforms?

![](_page_64_Picture_3.jpeg)

- **Data transfer and storage**: Moving large amounts of data to and from HPC systems can be time-consuming and costly. Additionally, storing data on HPC systems can be challenging due to limited storage capacity and the need for high-speed access.
- **Data management**: Managing and organizing large datasets on HPC systems can be difficult, as traditional file systems may not be well-suited for the high-performance needs of HPC applications.
- Data integrity and security: Ensuring the integrity and security of data on HPC systems can be challenging, as these systems are often shared by multiple users and may be exposed to a higher risk of data breaches.
- **Data visualization**: Visualizing and analyzing large, complex datasets on HPC systems can be difficult, as traditional visualization tools may not be able to handle the scale and complexity of the data.
- Scalability: Ensuring that data management solutions can scale to meet the needs of HPC systems as they grow in size and complexity can be a major challenge.

RDM-friendly architecture for HPC		
User Layer	Provides a flexible interface supporting <b>browsing</b> , <b>searching</b> , <b>depositing</b> , and <b>manipulating</b> data objects in the research data management system.	
Object Layer	Contains <u>data about the data</u> (e.g. <b>metadata</b> ), including where it is located in the storage layer.	
Storage Layer	Stores only data. The storage layer has to be a cost- effective, extendible mass storage system on which the	
Compute Ontario	Challenges of Research Data Management for High Performance Computing Schembera, Bjön : Brolisch, Thomas Research and Advanced Technology for Digital Libraries, 2017, p.140-151	

![](_page_65_Figure_1.jpeg)

![](_page_65_Figure_3.jpeg)

![](_page_66_Picture_1.jpeg)

Challenges of Research Data Management for High PerformanceComputingSchembera, Björn ; Bönisch, Thomas; Research andAdvanced Technology for Digital Libraries, 2017, p.140-151	Guide to writi Data Managen https://data.res
Like a rainbow in the dark: metadata annotation for HPC applications in the age of dark data. Schembera. (2021). The Journal of Supercomputing, 77(8), 8946–8966. https://doi.org/10.1007/s11227-020-03602-6	4 Tips for Met George Parke
Alliance storage and file management https://docs.alliancecan.ca/wiki/Storage_and_file_management	Zhang, Qian. Template: Adv <u>https://doi.org</u>
Alliance Learning and Training Resources	
training/training-resources	
Sharcnet Storage Structure and Policies	Other I
https://www.sharcnet.ca/my/systems/storage	
SciNot Data Managoment	UC San Diego
https://docs.scinet.utoronto.ca/index.php/Data_Management	https://researc
Storage 2020: A Vision for the Future of HPC Storage 2017-10-20 https://escholarship.org/uc/item/744479dp	Data-Intensive Intensive and Daniel C. M. d
Metadata Standards:	Data Manager
Texas Tech: https://guides.library.ttu.edu/c.php?g=/65394&p=569/025	Matthias Book
RDA: https://www.ucc.ac.uk/guluance/standards/metadata RDA: https://rdamsc.hath.ac.uk/	Mattinas Doei
io	

"readme" style metadata | Research t Service Group - Cornell University rch.cornell.edu/content/readme

ata Management IBC: https://osf.io/gsfc6

21). Data Management Plan ced Research Computing. Zenodo. .5281/zenodo.4573539

### adings

search Data Storage Explorer ta.ucsd.edu/finder

orkflow Management For Clouds and Datalable Computing Environments liveira, Ji Liu, and Esther Pacitti, 2019, p. 7

t in Machine Learning Systems Arun Kumar, and Jun Yang, 2019